確率微分方程式と統計的モデル評価 (統計的モデリングと予測理論のための統合的数理研究)

増田, 弘毅

増田, 弘毅. 確率微分方程式と統計的モデル評価 (統計的モデリングと予測理論のための統合的数理研究). 数理解析研究所講究録 2017, 2057: 81-89

数理解析研究所講究録
第2057巻 2017年 81-89

確率微分方程式と統計的モデル評価

九州大学大学院数理学府　増田　弘毅
Hiroki Masuda
Faculty of Mathematics, Kyushu University

本稿では, エルゴード的パラメトリック確率微分方程式 (SDE) モデルの相対評価規準に関する比較的最近の結果を, 背景にあるモデル評価規準の考え方とともに紹介する.

1　確率微分方程式モデルの相対評価規準

時系列データ $x_0, X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ が与えられているとする. それを Markov 型確率微分方程式 (SDE)

$$dX_t = A(X_t)dt + B(X_t)dw_t, \qquad X_0 = x_0, \tag{1.1}$$

の解過程 $X = (X_t)_{t\in\mathbb{R}_+}$ からの離散時点観測の実現値と捉えて統計モデルを構築し, データ生成機構を推定したい. ここで w は標準 Wiener 過程である. 基本設定として以下を考える.

- 離散時点観測データセット $\mathbf{X}_n = (X_{t_j})_{j=0}^n$, $t_j = jh_n$.
  - 観測間隔 $h_n > 0$ について, $n \to \infty$ のとき $nh_n^2 \to 0$ および $T_n := nh_n \to \infty$ と仮定する. これらはそれぞれ, 疑似尤度 (後述の (1.3)) を介してモデルを理論的に推定可能とするため, またモデルのエルゴード性を活用するためのものである. 実用上は h_n は数値としてユーザが指定する.
- エルゴード的パラメトリック SDE モデルの候補 M 個 $(m = 1, \ldots, M)$.

$$\begin{cases} dX_t = a_m(X_t, \alpha_m)dt + b_m(X_t, \beta_m)dw_t \\ X_0 = x_0 \end{cases} \tag{1.2}$$

  - 各モデル m において, ドリフト係数 a_m と拡散係数 b_m は未知パラメータは

$$\theta_m := (\alpha_m, \beta_m) \in \Theta_m \subset \mathbb{R}^{p_{\alpha_m}} \times \mathbb{R}^{p_{\beta_m}} = \mathbb{R}^{p_m}$$

    を除いて既知とし, またある値 $\theta_{m,0} = (\alpha_{m,0}, \beta_{m,0}) \in \Theta_m$ がとれて $a(\cdot, \alpha_{m,0}) = A(\cdot)$ かつ $b(\cdot, \beta_{m,0}) = B(\cdot)$ が成り立つ.

セミマルチンゲールの特性量の観点から, 各モデルでドリフト係数と拡散係数が決まれば X の分布が本質的に一意に定まる点に注意する [7].

単一のデータセットに対して係数モデル候補が複数ある状況であり, "M 個のうちどの候補モデルを採用すればよいか" をデータ駆動的に決めることを考えたい. ここでは情報量規準 (Information criterion, IC) に基づいたモデル選択の処方箋を紹介する. まずモデルを推定し, その結果を用いて選択方式を構成することになる.

各モデル m において未知パラメータがあるため, まず $\mathbf{X}_n$ に基づいて θ_m の正規型疑似推定量

$$\hat{\theta}_{m,n} = (\hat{\alpha}_{m,n}, \hat{\beta}_{m,n}) \in \operatorname{argmax} \mathbb{H}_{m,n}$$

でモデルの推定を行う. ここで

$$\begin{aligned}\mathbb{H}_{m,n}(\theta_m) = -\frac{1}{2}\sum_{j=1}^{n}\Bigg\{ &\log\det\left(2\pi h_n b_m^{\otimes 2}(X_{t_{j-1}}, \beta_m)\right) \\ &+ \frac{1}{h_n}(b_m^{\otimes 2})^{-1}(X_{t_{j-1}}, \beta_m)\left[\left(\Delta_j X - h_n a_m(X_{t_{j-1}}, \alpha_m)\right)^{\otimes 2}\right]\Bigg\},\end{aligned} \tag{1.3}$$

ただし $\Delta_j X = X_{t_j} - X_{t_{j-1}}$, $b_m^{\otimes 2}(x, \beta_m) = b_m b_m^{\top}(x, \beta_m)$, $A[u^{\otimes 2}] = u^{\top} A u$ である. この $\mathbb{H}_{m,n}$ は, θ_m が真値のときの Euler-丸山近似

$$X_{t_j} \approx X_{t_{j-1}} + a_m(X_{t_{j-1}}, \alpha_m) h_n + b_m(X_{t_{j-1}}, \beta_m)\Delta_j w$$

による遷移確率の正規近似

$$\mathcal{L}(X_{t_j} | X_{t_{j-1}} = x) \approx N\left(x + a_m(x, \alpha_m) h_n,\ B_m(x, \beta_m) h_n\right)$$

に基づいた疑似的な (微小時間近似型の) 尤度関数である. 実際 $\mathbb{H}_{m,n}$ は真の尤度関数ではないが, 明示的に書けることで数値的最適化が実行可能になるという多大な利点を持ち, さらに高頻度観測設定 $h_n \to 0$ が効いて理論的に最適 (漸近有効) であることが知られている. 特に漸近正規性

$$\left(\sqrt{T_n}(\hat{\alpha}_{m,n} - \alpha_{m,0}),\ \sqrt{n}(\hat{\beta}_{m,n} - \beta_{m,0})\right) \xrightarrow{\mathcal{L}} N_{p_{\alpha_m} + p_{\beta_m}}\left(0,\ \operatorname{diag}\{I_\alpha^{-1}(\theta_{m,0}),\ I_\beta^{-1}(\theta_{m,0})\}\right) \tag{1.4}$$

が成り立ち, これより, 想定された真のモデル (1.1) に対する標準的な信頼領域の構成や統計的仮説検定の定式化が自然に考察可能となる ([8], [9], [15]).

以上の下, IC として以下が知られている.

CIC (Contrast based Information Criterion, [13])

$$\mathrm{CIC}_{m,n} = -2\mathbb{H}_{m,n}(\hat{\theta}_{m,n}) + 2p_m \tag{1.5}$$

QBIC (Quasi-Bayesian Information Criterion, [6])

$$\begin{aligned}\mathrm{QBIC}_{m,n} = &-2\mathbb{H}_{m,n}(\hat{\theta}_{m,n}) \\ &+ \log\det\left(-\partial_{\alpha_m}^2 \mathbb{H}_{m,n}(\hat{\theta}_{m,n})\right) + \log\det\left(-\partial_{\beta_m}^2 \mathbb{H}_{m,n}(\hat{\theta}_{m,n})\right)\end{aligned} \tag{1.6}$$

ここで ∂_a は変数 a に関する偏微分を表す.

BIC (Bayesian Information Criterion, [6])

$$\mathrm{BIC}_{m,n} = -2\mathbb{H}_{m,n}(\hat{\theta}_{m,n}) + p_{\alpha_m}\log T_n + p_{\beta_m}\log n \tag{1.7}$$

CIC は Akaike information criterion (AIC) [1, 2] に, また BIC と QBIC は Schwarz の BIC [12] に対応した IC である. モデルには係数関数の滑らかさ・非退化性やエルゴード性などの正則条件が必要であり, このような理論背景に裏打ちされることで IC の合理性が確保される (詳細は上記各 IC の参考文献を参照).

各 IC において, M 個の値 $\mathrm{IC}_{1,n},\dots,\mathrm{IC}_{M,n}$ をデータから計算し, 最小となるモデル m を最適モデルとして採用する. ユーザにとって留意すべきは "最適" の意味である. IC の種類が異なれば当然最適モデルも異なり得る.

2 統計モデルの相対評価

前節の IC は Kullback-Leibler 情報量を介して統一的に導出できる. 本節ではこの背景をやや一般的な枠組みで概観していく. より詳細については, 和書では [18], [17], [19], 洋書では [3], [4] などが参考になる. 先駆者による回顧的な記事 [16] も挙げておく (ただし [16] では BIC と Kullback-Leibler 情報量を明示的にリンクさせていない). また, 検定との絡みを軸に AIC と BIC を非形式的に比較した読み物として [5] がある.

2.1 基本的な考え方

見やすさのため記号を改めて設定する. 観測データを確率変数 X とし, その確率分布を

$$\mathbb{P}(X \in dx) = g(x)\mu(dx)$$

で表す. ここで μ は適当な参照測度である (Lebesgue 測度や計数測度が典型的). 未知の μ-密度 g に興味があるが, データ X に不確実性が伴う中では正確には知り得ないため, X の実現値に基づいてこれを推定したい.

分析者が g のための複数の統計モデル候補 $f_1,\dots,f_M$ を用意してそこから一つを選びたい場合, 先述の通り最適性指標の設定がポイントとなる. SDE モデルの場合, (1.1) で定まる X から得られる離散時点データの真の分布が g に, 統計モデル (1.2) を推定するために仮想的に想定した正規型疑似尤度が f_m に対応する. 候補モデル $f_1,\dots,f_M$ の設定はユーザ依存であるため, 統計的モデル評価は相対的なものである. 実際, f_m が全て g からかけ離れた場合にも定量的に相対評価が可能である. 留意すべき点としては,

- 仮想的な統計モデル f_m は "データから推定可能" でなければならないこと,
- 異なる最適指標が理論上相反することもあり, 絶対的な評価規準は無いこと

などが挙げられる. 評価の方法論の主たるものは AIC と BIC であり, 提唱後数十年が経っていた現在でも, それらは統計的モデル評価規準における双璧であり続けている. AIC と BIC は見た目は似ているが数理的な性質を異にする.

2.2 AIC: 赤池の思想

モデル評価においては，候補モデル f_m と真のモデル g の間の"乖離度"の定量的指標として何を採用するかが重要である．分布 g から見た f_m への Kullback-Leibler 情報量

$$D(f_m;g) := \mathbb{E}\left(\log\frac{g(X)}{f_m(X)}\right) = \int\left(\log\frac{g(x)}{f_m(x)}\right)g(x)\mu(dx) \tag{2.1}$$

を考えよう．対称性がない $(D(f_m;g) \neq D(g;f_m))$ 点に注意する．Jensen の不等式より常に $D(f_m;g) \geq 0$ であり，f_m と g が μ-a.e. で一致するときに限って $D(f_m;g)=0$ となる．

特に各 f_m を手元のデータ X からパラメトリックに推定する場合，それらを明示的に $f_m(\cdot;\theta_m)$ と書くとき，適当な推定量 $\hat{\theta}_{m,n}=\hat{\theta}_{m,n}(X)$ で推定されたモデル（予測モデル）$\hat{f}_m$ は

$$\hat{f}_m(x) = f_m(x;\hat{\theta}_{m,n}) \tag{2.2}$$

の形になる．候補として当てがうモデル $f_m(x;\theta_m)$ は，如何なる θ_m に対しても $g(x)$ と一致する必要はない．しかし，SDE モデル (1.2) に対する正規型疑似尤度のように真の尤度ではない疑似尤度を用いる場合でも，(1.4) のように真のモデル構造を推定可能な状況は少なからず存在する．

"偶然そうなっただけ"という解釈を避けるため，統計的意思決定の背後にある理論ではランダムでない何らかの"期待値量"で定まる指標を据えるのが基本である．今 Kullback-Leibler 情報量 (2.1) を最小にするようなモデル候補を選びたい．(2.2) の形を考える場合には，推定量 $f_m(x;\hat{\theta}_{m,n})$ を代入した期待値

$$-\mathbb{E}'\otimes\mathbb{E}\left\{\log f_m\big(X;\hat{\theta}_{m,n}(X')\big)\right\} := -\iint \log f_m\big(x;\hat{\theta}_{m,n}(x')\big)g(x)\mu(dx)g(x')\mu(dx')$$

を最小にするモデルを選択することになる．しかしこれは統計量でないため，何らかの形で推定する必要が生じる．単純には $-\log f_m\big(X;\hat{\theta}_{m,n}(X)\big)$ を用いたいところであるが，適当な統計量 $\hat{b}_{m,n}(X)$ をもって，

$$\begin{aligned}
&-\mathbb{E}'\otimes\mathbb{E}\left\{\log f_m\big(X;\hat{\theta}_{m,n}(X')\big)\right\} - \left(\mathbb{E}'\otimes\mathbb{E}\left\{-\log f_m\big(X';\hat{\theta}_{m,n}(X')\big)+\hat{b}_{m,n}(X')\right\}\right)\\
&=\int\left\{\left(-\int\log f_m\big(x;\hat{\theta}_{m,n}(x')\big)g(x)\mu(dx)\right)\right.\\
&\qquad\left.-\left(-\log f_m\big(x';\hat{\theta}_{m,n}(x')\big)+\hat{b}_{m,n}(x')\right)\right\}g(x')\mu(dx') = o(1), \quad n\to\infty,
\end{aligned}$$

なる形で近似的なバイアス補正を行うことを考える．尤度原理で理詰めて最後にバイアス補正，すなわち期待値量に着目することで統計量構成を図る，これが赤池の思想の本質と言える．このとき統計量 $-\log f_m\big(X;\hat{\theta}_{m,n}(X)\big)+\hat{b}_{m,n}(X)$ を IC として用いることが可能となる．慣例により，"2"を乗じた

$$-2\log f_m\big(X;\hat{\theta}_{m,n}(X)\big)+2\hat{b}_{m,n}(X)$$

なる形で定義される．

特に AIC [1, 2] は補正項が $\hat{b}_{m,n}(X) = p_m$ (パラメータの次元) で与えられる場合に該当し,

$$\mathrm{AIC}_{m,n} = -2\log f_m\big(X; \hat{\theta}_{m,n}(X)\big) + 2p_m \tag{2.3}$$

で与えられる. 構成から明らかなように, AIC 型の IC は Kullback-Leibler 情報量予測誤差の期待値量を近似不偏的に最小にするモデル選択規準である. AIC は非常に簡潔であるがその意味するところは深く, 赤池の思想は統計的モデリングに多大な影響をもたらした. 今日 AIC の変形版やその考え方を一般化して得られた情報量規準は数多く存在するが, モデルの予測力ベース (汎化誤差の最小化) のものの大半は赤池の思想に基礎を据えていると思われる. AIC によって選択されたモデルは, 一種の予測力の意味で最適となる (e.g. [3], [4]). 赤池の偉業はモデル複雑度と適合度をバランスをベースとした統計的モデリング手法の哲学創成ともいうべき点に宿る. AIC の業績を讃えられ, 2006 年に京都賞基礎科学部門を受賞された.

2.3 BIC: Schwarz の Bayes 型モデル評価

BIC は Gideon Schwarz [12] により導入されたモデル評価規準で, SIC ともよばれる. BIC は, 真のモデル g から推定されたモデルへの乖離度を Kullback-Leibler 情報量で測って最適なものを採用するという赤池のロジックに基づいて導出可能である. 70 年代後半に "AIC vs BIC" の論争が勃発したが, そもそも AIC と BIC は目的に応じて使い分けられるべき手法であるのが事実である [3], [4], [5], [16], [17].

BIC は Bayes モデルをベースとして構築され, パラメータ θ を確率変数と捉える点において先述の最尤推定法と決定的に異なる. 基本的な構成要素はパラメータの事前分布 $\pi(\theta)$, $\theta \in \Theta \subset \mathbb{R}^p$, と θ 所与でのデータ分布 $f(x;\theta)$ である. 今 M 個の Bayes モデル候補

$$\{(\pi_m(\theta_m),\, f_m(x;\theta_m));\ \theta_m \in \Theta_m \subset \mathbb{R}^{p_m}\}, \qquad m = 1,\ldots,M,$$

が与えられているとしよう. パラメータ θ_m の事前分布 $\pi_m(\theta_m)$ は通常恣意的に決められる. 観測 x において, モデル m における周辺尤度は

$$f_m(x) := \int_{\Theta_m} f_m(x;\theta_m)\pi_m(\theta_m)d\theta_m \tag{2.4}$$

で定義される $(m = 1,\ldots,M)$. 周辺尤度はモデルエビデンスともよばれ, データとパラメータの同時尤度をデータへ周辺化したものを指す. AIC は予測モデル $\hat{f}_m(\cdot) = f_m(\cdot;\hat{\theta}_{m,n})$ に関する Kullback-Leibler 情報量最小化によって得られたが, BIC は周辺尤度 $\hat{f}_m(\cdot) = f_m(\cdot)$ を最大化し, 実際のデータを与える確率・頻度が最も高いモデルを選択することで得られる. 周辺尤度 (2.4) の積分計算を陽に行うことは一般に困難なため, 何らかの近似的手段に訴えたい.

ある正定値定数行列 $|A_{m,n}| \to 0$ $(n \to \infty)$ に対して $A_{m,n}^{-1}(\hat{\theta}_{m,n} - \theta_{m,0})$ が漸近混合正規性を持つなど, 適切な正則条件の下で周辺尤度は以下のように近似できる (今扱っている SDE モデルの場合を含む).

$$\log f_m(X) = \log f_m(X;\hat{\theta}_{m,n}) - \frac{1}{2}\log\det\Big(-\partial_{\theta_m}^2 \log f_m(X;\hat{\theta}_{m,n})\Big)$$

$$+ \log \pi_m(\hat{\theta}_{m,n}) + \frac{1}{2} p_m \log 2\pi + o(1). \tag{2.5}$$

主要項として確率的に発散する項を束ねたもののみを取り出して “-2” を乗じることで, 以下の IC が得られる.

$$\mathrm{QBIC}_{m,n} = -2 \log f_m(X; \hat{\theta}_{m,n}) + \log \det \left(-\partial^2_{\theta_m} \log f_m(X; \hat{\theta}_{m,n}) \right). \tag{2.6}$$

通常 (2.5) は

$$\log f_m(X) = \log f_m(X; \hat{\theta}_{m,n}) - \frac{1}{2} \log |A^{-2}_{m,n}| + O(1) \tag{2.7}$$

と書き換えることができる. このため $\mathrm{QBIC}_{m,n}$ と漸近同等な

$$\mathrm{BIC}_{m,n} = -2 \log f_m(X; \hat{\theta}_{m,n}) + \log |A^{-2}_{m,n}| \tag{2.8}$$

を用いることも考えられるが, 当然, 有限標本においては QBIC と BIC によるモデル選択結果は異なり得る. なお, 適当な条件の下, (2.5) と (2.7) におけるオーダー記号はそれぞれ L^q ($q > 0$ は任意) の意味で成立する [6].

元々BIC [12] は Kullback-Leibler 情報量最小化ではなく各候補モデルの生起事前確率を設定して Bayes の定理からそれぞれの生起事後確率, つまり恣意的に導入した事前分布を観測データで更新した分布を計算し, それが最大のモデルを選択するというシナリオで導出された. Bayes 統計的観点からはこちらの方が捉えやすいというか自然に見えるかもしれない. このように “モデル生起確率” なる設定を許容した下で, モデル生起事前確率を設定すれば, データを観測した後のモデル m の生起事後確率は BIC 統計量を介して

$$\frac{\exp(-\mathrm{BIC}_{m,n})}{\sum_{l=1}^{M} \exp(-\mathrm{BIC}_{l,n})}, \qquad m = 1, \ldots, M,$$

と近似できるため, Bayes 型モデル評価基準では, “各候補モデルの生起事前確率および各モデルにおけるパラメータの事前分布の設定の必要性” と引き換えに, “定量的なモデル選択の事後評価事後評価” が自然に可能となる.

BIC 以外にも様々な Bayes 型モデル評価基準が存在する. 特に MCMC などを軸に, 計算機を駆使したモデル評価を行う手法も多々存在する. 何にせよ, 重要なのは合理性を持って適用可能な手法かどうかであり, 個々の規準の意味の正しい理解は不可欠である.

2.4 確率微分方程式モデル再訪

前節で触れた (2.3), (2.6), (2.8) をそれぞれ SDE の文脈で導出することで, (1.5), (1.6), (1.7) が得られる. いずれも偽りの疑似尤度を用いた方式であるが, 理論上好ましい性質を有する.

- CIC (1.5) は, 実は正体不明な真の尤度に正規型疑似最尤推定量を代入したものと漸近不偏性の意味で同値となる [13]. 同論文では, 常に大きめのモデルを選択するという AIC の傾向が数値実験で観察されているため, 各モデルの選択確率が $n \to \infty$ のとき理論上どう落ち着くかも解析可能であろう.

- QBIC (1.6) および BIC (1.7) の導出に際して, [6] は [15] の多項式型大偏差確率の評価式を適用した. ともによく知られた BIC 型統計量の利点である "モデル選択の一致性" を持つ. すなわち, 真のモデル (より広くは Kullback-Leibler 情報量の意味で最適なモデル) を $n \to \infty$ で 1 へ近づく確率で選択する [6]. なお, (Q)BIC 型のロジックにおける疑似尤度と真の尤度とのリンクは未調査の状態である. [6] における階層型 SDE モデルに対する数値実験により, BIC は比較的小さめモデルを選ぶという典型的な傾向が観察されている. また QBIC は総じて CIC と BIC の中間的な振る舞いを呈するが, 本稿では触れていない非エルゴード的確率過程モデルの場合では, QBIC が BIC のパフォーマンス (真のモデルの選択頻度) を優越する現象が見られた.

3 補遺

特にデータを生成する物理モデルが未知または存在しない状況において定量的かつ合理的な意思決定を行うためには統計解析は必要不可欠であり, とりわけ統計モデルの構築ならびに相対評価はその中枢を成す. 今日まで様々なモデルの相対評価手法が提案されてきているが, それらの根底には AIC もしくは BIC の思想が根強く脈打っている.

3.1 一般事項

1. AIC のロジックを数学的に拡張・精密化することは多方向へ考えられる. 例えば, 推定量は基本的に何であっても (バイアス $\hat{b}_{m,n}(X) = p_m$ は変わり得るが) その理念は変わらないし, 分布間の乖離度を測る Kullback-Leibler 情報量をより一般のダイバージェンスに変えたモデル評価規準の定式化も可能であろう. 特に Kullback-Leibler 情報量は異常データに対して脆く, たった一つの異常値があるだけでも結果に大きく響く場合が多い. このため, 異常値データに頑健な別のダイバージェンスに対して AIC のロジックを踏襲することで, 意味のある情報量規準の構築が期待できる. 具体的な試みは [10] などでなされている.

2. 複数あるモデル候補から単一のものを選択するのではなく, ある程度信頼出来るモデルの集合を見出す "モデル信頼集合" なる概念もある. さらには各モデルでの推定結果を融合し, 力を併せて予測問題に役立てようという "モデル平均化法" も提案されて以来 15 年以上になる ([4] とその参考文献参照). モデル選択の不確実性を加味し, 危険性を緩和しつつ予測能力の向上も図ろうという方法論である.

3. モデルインデックスが $m = 1, \ldots, M$ と離散的でなく連続変量で表現される場合も IC の考え方は全く同様である. とりわけ Bayes 的視点でパラメータに事前分布を入れた正則化推定法は, パラメータの次元 (例えば説明変数の次元) の大きさに起因する推定の不安定さへの対処法として近年一つの標準的方法論として確立するに至った. 正則化推定は, パラメータの何らかのノルムを抑える事前分布を取り込んで推定を行うことに相当する. 特にスパース型正則化推定については, 最近のサーベイ

和文誌 [20] とその参考文献に多くの情報がある. 最近 [11] により, スパース推定の枠組みで AIC を導出する試みもなされた.

4. 歴史的に AIC は, 全候補モデルが真のモデルを含んでいる場合において定式化された. AIC を適用するにあたっては候補モデルはどれもそれなりに大きく設定される (複雑である) べき, ということである. とは言え, モデル候補が誤特定, すなわち真のモデルを含んでいない場合においても IC 導出はパラレルに展開される.

3.2 確率微分方程式モデル関連事項

1. 観測頻度を表す h_n はユーザ側で具体的数値を設定する必要があった. 小さくないと漸近論が機能しないことに注意しつつ, 実用上は T_n とのバランスを加味して値を決めることになる. 確率過程統計モデルにおける時間スケールと実時間スケールの対応を, 何らかの指標をもって統計的に設定できれば有意義であろう.

2. Lévy 過程で駆動されるジャンプ型 SDE モデルの相対評価理論・方法論については, 一般的な結果は得られていない.

3. 係数の関数形が誤特定された SDE モデルの場合, 漸近正規性 (1.4) が本質的に変わる [14]. この場合の CIC と (Q)BIC の性質については明らかにはされていない.

4. シミュレーション, 漸近展開公式, 統計推測など, SDE モデルに係る様々な解析ツールが統計解析 R の YUIMA パッケージへ実装されている.

   `https://r-forge.r-project.org/projects/yuima/`
   `https://cran.r-project.org/web/packages/yuima/index.html`

   YUIMA は随時更新されており, SDE モデル評価も実装予定である.

参考文献

[1] H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Tsahkadsor, 1971)*, pages 267–281. Akadémiai Kiadó, Budapest, 1973.

[2] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automatic Control*, AC-19:716–723, 1974. System identification and time-series analysis.

[3] K. P. Burnham and D. R. Anderson. *Model selection and multimodel inference.* Springer-Verlag, New York, second edition, 2002. A practical information-theoretic approach.

[4] G. Claeskens and N. L. Hjort. *Model Selection and Model Averaging.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2008.

[5] J. J. Dziak, D. L. Coffman, S. T. Lanze, and R. Li. Sensitivity and specificity of information criteria. *PeerJ PrePrints*, 3, 2012.

[6] S. Eguchi and H. Masuda. Schwarz type model comparison for LAQ models. *arXiv:1606.01627*, 2016.

[7] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes.* Springer-Verlag, Berlin, second edition, 2003.

[8] M. Kessler. Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, 24(2):211–229, 1997.

[9] H. Kitagawa and M. Uchida. Adaptive test statistics for ergodic diffusion processes sampled at discrete times. *J. Statist. Plann. Inference*, 150:84–110, 2014.

[10] K. Mattheou, S. Lee, and A. Karagrigoriou. A model selection criterion based on the BHHJ measure of divergence. *J. Statist. Plann. Inference*, 139(2):228–235, 2009.

[11] Y. Ninomiya and S. Kawano. AIC for the lasso in generalized linear models. *Electron. J. Statist.*, 10(2):2537–2560, 2016.

[12] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464, 1978.

[13] M. Uchida. Contrast-based information criterion for ergodic diffusion processes from discrete observations. *Ann. Inst. Statist. Math.*, 62(1):161–187, 2010.

[14] M. Uchida and N. Yoshida. Estimation for misspecified ergodic diffusion processes from discrete observations. *ESAIM Probab. Stat.*, 15:270–290, 2011.

[15] N. Yoshida. Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Ann. Inst. Statist. Math.*, 63(3):431–479, 2011.

[16] 赤池 弘次. AIC と MDL と BIC. オペレーションズ・リサーチ: 経営の科学 41(7), 375–378, 1996-07-01, 1996.

[17] 赤池弘次, 甘利俊一, 北川源四郎, 樺島 祥介, 下平 英寿. 赤池情報量規準 AIC–モデリング・予測・知識発見. 共立出版, 2007.

[18] 小西 貞則, 北川 源四郎. 情報量規準. 朝倉書店, 2004.

[19] 下平 英寿, 伊藤 秀一, 久保川 達也, 竹内 啓. モデル選択–予測・検定・推定の交差点. 統計科学のフロンティア 3, 岩波書店, 2004.

[20] 廣瀬慧. スパースモデリングとモデル選択. 電子情報通信学会誌, 99 巻, 5 号, 392–399, 2016.